## *Introduction*

Recent school science curriculum frameworks around the world have reflected science as more than a body of facts (see, for instance, Australian Curriculum Assessment and Reporting Authority 2009; Duschl, Schweingruber, and Shouse 2006; Ekiz 2006; Ministry of Education, Culture and Science, Netherlands undated; Ministry of National Education, Science and Technology Curriculum, Turkey 2005; Qualifications and Curriculum Authority 2004). Pupils are now expected also to engage with the central role of evidence in science. In science the traditional curriculum emphasis in classroom teaching has largely been on the familiar factual knowledge, concepts, laws and theories of science. This is traditionally known as substantive or conceptual understanding. The facts, concepts, laws and theories that contribute to this substantive understanding are, of course, themselves supported by empirical evidence or are subject to investigation. Traditional substantive understanding alone is not sufficient to describe the ideas of science. Another key curriculum component is concerned with the procedures of science.

The UK's national academy of science, the Royal Society, has the motto *Nullius in verba* which, roughly translated, means 'take nobody's word for it'. This phrase encapsulates the defining characteristic of science; that is that it is based on evidence, not opinion. An understanding of evidence and knowing how it can be applied to solve problems and to evaluate scientific claims is arguably a key element of scientific literacy (Gott and Duggan 2007), which is an aim of these recent curriculum developments.

Since pupils around the world are now expected to engage with the central role of evidence in science, teacher education programmes need to prepare pre-service teachers (henceforth referred to as 'student teachers') to teach about evidence. An understanding of evidence in science comprises part of what Shulman (1986, 1987) termed the Content Knowledge for teaching and Grossman (1990) identified as Subject-Matter Knowledge.

We have researched (Roberts and Gott 2007, 2010; Glaesser et al. 2009a, 2009b; Roberts, Gott, and Glaesser 2009) student teachers' understanding and application of evidence before and after a module (termed the Evidence Module) that explicitly teaches the 'concepts of evidence' (Gott, Duggan, and Roberts 2003) in England. Given the constraints of time that we face in our pre-service programmes, we have found that the short module is an efficient and effective way of developing student teachers' content knowledge about evidence: specifically their understanding of the ideas of evidence; their ability to carry out an open-ended investigation; and their ability to ask questions that focus on the evidence behind scientists' claims in the context of a socio-scientific issue. In this research we report a replication study with student teachers in Turkey.

The importance of the replication study is that the student teachers in each country had different prior experiences, as pupils, of school science and therefore may have started the module with different backgrounds. Could the same module develop an understanding about evidence despite these differences? In England, student teachers' background of National Curriculum science (DES and Welsh Office 1989) included a procedural component entitled Scientific Enquiry which comprised one of four strands of the curriculum. Pupils' reports of practical investigations were assessed in GCSEs (taken at age 16, at the end of compulsory schooling) and practical work was, and still is, an important element of science teaching in English schools. However in contrast to this, as school pupils, student teachers in Turkey followed a more traditional curriculum which focused on the substantive ideas of science and included very little practical work (Demirel 2004; Ministry of National Education, Science Curriculum 2000; Sahin-Pekmez 2000, 2005; Unal, Costu, and Karatas 2004). Yet, as future teachers in England and Turkey, both cohorts need to develop their subject-matter or content knowledge about evidence since both countries' school science curricula emphasise procedural understanding (Ministry of National Education, Science and Technology Curriculum, Turkey 2005; Qualifications and Curriculum Authority 2004). Would the Evidence Module work in Turkey?

## *Teaching the procedural component of the science curriculum*

The procedural component of the curriculum is concerned with 'doing science' (Hodson 1991). Polanyi (1966) considered some forms of expert procedural knowledge, 'knowhow', to be tacit and unable to be codified. From this perspective, expertise is more than the sum of the component parts. But what, at least, are these component parts? What do we need to teach students to help them to investigate and understand evidence?

The different perspectives of how the procedural component is conceived can be characterised by examining two contrasting perspectives: a 'skills' perspective and an 'understanding of ideas about evidence' perspective. While we see these as two fundamentally different perspectives we recognise that some literature contains some elements from both. The perspectives taken will, in turn, influence both how the procedural component is taught and assessed. We will attempt to illustrate this in the following account.

### The Skills Perspective

The skills perspective is characterised by performance of 'process skills'. The main characteristic of such a perspective is that the procedural component is perceived as a set of skills that can be practiced and are learned by *repeated* exposure to practical work, which obviously takes time. The procedural component is largely implicit in teaching and any guidance given to students is through a simple exemplification of the processes. The prior school experience of the English students in this research was largely from this perspective (Roberts and Gott 2003) while the Turkish students had experienced little practical work.

The skills perspective has a long history. An early version is typified by the Science: A Process Approach developed from work by Gagné which identified isolated 'process skills' (American Association for the Advancement of Science 1967) which was followed by others including, in the UK, Warwick Process Science (Screen 1986) which emphasised 'process skills' such as observing, classifying and interpreting. Paralleling the developments to teach process skills were assessment schemes such as The Assessment of Practical Science (TAPS; Bryce et al. 1983) and the Graded Assessment in Science Project (GASP; Davis 1989) which were both based on the assessment of 'process' in the context of performance of isolated practical skills.

Research has shown that 'children failed to develop meaningful understanding under science-as-process instructional programs … but its legacy persists in both policy and practice' (Duschl, Schweingruber, and Shouse 2006, 8: 2–3). Elements of that legacy can still be seen in curricula that either have procedural components specified as behavioural objectives, since these may be translated into classroom practice and assessment as just 'doing', or in curricula that emphasise using investigations as a pedagogical approach, a way of teaching, for mainly substantive understanding. In such pedagogical approaches the 'doing' of science is considered to be sufficient to meet the procedural component of the curriculum; students 'discover' the procedural element with practice.

### An Understanding of Ideas about Evidence Perspective

An understanding of ideas about evidence perspective is a different way of conceptualising the procedural component of the curriculum. The procedural component is seen to be underpinned by a set of ideas about evidence. It is constructivist at heart since it requires the learner to construct meaning, specifically about the concepts of validity and reliability, from specific ideas about evidence. The focus is on a set of ideas that are an integral part of science and that can then be learned, understood and applied, rather than a set of skills that are thought to develop implicitly by practice.

These ideas can be applied and synthesised in open-ended investigations, together, of course, with the traditional substantive ideas of science. We also consider them to be important in empowered forms of scientific literacy, to enable students to engage with scientists' claims and scientific argumentation (Gott and Duggan 2007; Gott and Roberts 2008; Tytler, Duggan, and Gott 2001a, 2001b). As Buffler, Allie and Lubben (2001) argue: 'Procedural knowledge (in the context of experimental work) will inform decisions, for example, when planning experimental investigations, processing data and using data to support conclusions' (p.1137).

We have referred to these ideas about evidence as 'the thinking behind the doing' (not to be confused with meta-cognitive notions of 'thinking about one's own thinking') and have created a tentative list numbering some 80 or so of them which we have called the 'concepts of evidence' (Gott, Duggan, and Roberts 2003). They serve, we argue, as a domain specification of ideas necessary for procedural understanding. From this perspective, the procedural component of a curriculum consists of ideas that form a knowledge-base of evidence that can be explicitly taught and assessed, in a similar way to the more traditional substantive elements in the curriculum, using both practical and non-practical activities. It is this perspective that underpins the rationale for the Evidence Module and instruments used in this research.

The concepts of evidence include ideas about the uncertainty of data (as taught and researched by Buffler, Allie, and Lubben 2001). They also include ideas important to understanding measurement and data processing, presentation and analysis which may be considered to be part of the mathematics curriculum but which are essential for understanding evidence. It is therefore reasonable to assume that both our English and Turkish students, regardless of the differences in their previous experiences with practical work, will have elements of such an understanding from their mathematics curriculum.

The concepts of evidence are a toolkit of ideas that can be understood in their own right but are integral to the planning and carrying out of practical investigations with understanding (rather than as a routinised procedure (Roberts and Gott 2003)) and the evaluation of others' evidence.

In this research we utilise instruments that measure student teachers' understanding of the ideas of evidence; their application in an investigation; and their role in evaluating a claim in a socio-scientific context. We will now describe how the 'understanding ideas about evidence' perspective has been used in this research.

## The Concepts of Evidence (CofEv)

In this section we explain the descriptive models that structure the research design and influence the coding of the instruments. They are described more fully elsewhere and readers are referred to the original papers for further details (Roberts and Gott 2007, 2010; Glaesser et al. 2009a, 2009b; Roberts, Gott, and Glaesser 2009).

The extensive list of the concepts of evidence (Gott, Duggan, and Roberts 2003) was the basis of the ideas taught in the Evidence Module. The list is, however, unwieldy for some analysis purposes (but critical for curriculum definition purposes). We have grouped the ideas together into subsets, or layers, represented in the 'bull's-eye' diagram (Figure 1) which we expand on now so that the reader can understand more about the 'understanding ideas about evidence' perspective that underpins this research.

[Insert FIG 1 ABOUT HERE]

***Layer A: A single datum***

In any empirical investigation something will need to be measured, either qualitatively or quantitatively. This innermost layer is to do with the making of a single measurement of that variable. It is at the heart of science and hence is central in the bull's-eye. The ideas in this layer must take into account, inter alia, the range and sensitivity of the instrument, the validity of the measurement and its accuracy. If the measurement is invalid or unreliable then the validity of the whole investigation and any claim made from it is called into question. But, of course, one measurement is not always enough and repeats will be necessary.

***Layer B: A data set***

We are led, then, into the next layer which includes ideas to do with repeated measurements of the same variable under the same conditions: whether sufficient repeated readings have been taken to capture the variation - whether that is variation in the sample or variation inherent in the measurements themselves.

***Layer C: Design and Relationships***

As we move out, we arrive at the crux of the investigation; the establishing, or otherwise, of a relationship between one or more of the variables. This layer really permeates the whole task, but it is hard to show that in a simple diagram. When seeking relationships between variables - whether in controlled lab-based conditions or in surveys of naturally changing variables - the validity of the design must be considered as well as how the reliability of the data affects the interpretation of the relationship. For instance, decisions will need to be made about the range of the independent variable which will be needed to establish any potential relationship as well as the interval between such readings and also how potentially confounding variables are treated.

These 3 inner layers (A-C) represent the core concepts of evidence in an investigation. These inner layers, in turn may have been influenced by ideas subsumed in the outer 3 layers, D - F: ideas from other similar research, which must be evaluated; the expertise of the investigators themselves and how this might affect the quality of the evidence being collected; and also any potential economic and social pressures influencing the design and conduct of the investigation. These ideas could potentially bias any stage of the data collection. An evaluation of the validity and reliability of the whole investigation should take account of all of these potential influences.

These ideas are taught in the Evidence Module and are assessed explicitly in the Evidence Test used in this research.

They also are the ideas that an investigator will draw on to make decisions in an investigation. The aim of the investigation is to make a claim arguing, as it were, from the centre of the bull's-eye and working outward from the data. We represent this in Figure 2 and have described it as 'looking forward' (Gott and Duggan 2003). This is not to imply that the investigator works in a linear fashion; making decisions in the light of emerging evidence during an investigation (which we have described as working iteratively (Roberts, Gott, and Glaesser 2009) is a sophisticated approach to the problem and a creative use of these ideas. The arrow represents how the investigator aims to ensure the quality of each and every datum, and will consider the validity and reliability of the evidence in each layer that may have impacted on the data so they can make a qualified claim or conclusion. In our coding for the Open-ended Investigation we consider both the ideas students have used as well as their approach to the investigation.

Many scientific claims are related to issues in the public domain. This is the point at which decisions are taken and where the scientifically literate person is likely to become involved. In such a context, the scientifically literate person will not have been directly involved with the investigation but may be presented with a scientist's claim from empirical work. By 'looking back' from the claim they can, if they so wish, question and evaluate the evidence for the claim using the same ideas. In effect, they need to put themselves in the position of the investigator and question the 'thinking behind the doing' that must have occurred. This is represented on Figure 2.

[INSERT FIG 2 ABOUT HERE]

But many such socio-scientific issues are complex; decisions need to be made in a context where the scientist's claim will be just one element of a 'broader debate' in which the scientist's claim is just one of the factors to be considered (Roberts 2009). The decision will often be influenced by other, possibly competing, scientific claims on the same issue. Added to that will be 'value' claims linked to ethics, or aesthetics, or politics, and empirically based economic claims, to say nothing of totally unscientific and even irrational claims based on prejudice and ignorance. We have argued that in such a context a scientifically literate person may want to interrogate the scientist's claim, to determine its weight, by evaluating the validity and reliability of the evidence that lead to the claim.

The third instrument used in this research, the Issue Task, analysed whether student teachers, when presented with the scientist's claim in relation to a socio-scientific issue engaged in the broader debate; and whether they asked questions about the evidence for a claim, by 'looking back' to the investigation that lead to it.

We can now see how the concepts of evidence underpin the teaching and the instruments in this work.

## The research

This research, carried out in both countries, compares cohorts of student teachers' understanding about evidence before and after they have been explicitly taught procedural ideas. Gott and Roberts (2008) have previously carried out research with primary teacher education students in England and this paper reports a replication study with Turkish primary teacher education students.

The main question in this study is:

- Can the Evidence Module develop understanding about scientific evidence and its application in both cohorts of students?

Our work is small scale and exploratory. We attempt to explore whether a module that teaches the concepts of evidence and includes open-ended investigation enables these cohorts of students to understand and apply ideas about evidence. We are not suggesting that these cohorts are typical of others in England or Turkey.

We undertook this research without any great expectations of success since the Evidence Module was developed in England over many years and we were wary about how it might transfer to a new context. But, if it did turn out to be effective, it would mean that we had a parsimonious way of teaching this subject-matter or content knowledge to pre-service teachers.

## Method

In summary, the research in each country took the form of an intervention, the Evidence Module, with pre-tests and post-tests using the same three instruments.

In England the teaching and testing was carried out in English. In Turkey the student teachers are sometimes taught using English resources. The original English teaching materials were supplemented with Turkish translations and Turkish colleagues provided further explanations in Turkish. The Evidence Test was presented in English; the Open-ended Investigation and Issue Task were in Turkish. All responses to instruments were in Turkish.

**The sample**

The sample consists of 38 Turkish student teachers and 91 English, both on the 2nd year of BA primary teaching courses. We do not have both pre- and post- data from the whole sample in the English cohort, and thus we must be cautious in any analysis and claims.

The student teachers agreed to participate with the research which was approved by the relevant university ethics committees.

The sample is described here (Table 1) so that readers can locate the sample in comparison to students in their own experience. We collected biographical data from student questionnaires and the university academic records. They are mainly female in both countries but there are significant differences in their school science background, including their previous experience of practical work (as described above) and their previous attainment. The Turkish students have taken science at a higher level than the majority of English students and appear to have much higher qualifications on entry to university (at least when compared to other programmes in the same universities). We recognise that these differences may significantly affect the student teachers' pre-test responses and the effect of the intervention. The purpose of the research is to replicate research carried out in England, to determine if the Evidence Module can help to develop student teachers' content knowledge about scientific evidence despite these many differences.

[INSERT TABLE 1 ABOUT HERE]

**The Evidence Module**

The Evidence Module is described in detail in Gott and Roberts (2008).

Exactly the same teaching materials and background resources were used in England and Turkey. They were presented to the student teachers in English. In Turkey, some projected presentations had additional notes in Turkish and the lecturers taught in both English and Turkish, as was their usual practice.

In summary, the module for each cohort lasted a term. Teaching consisted of both lectures and workshop activities. Practical and non-practical approaches were used in the teaching. The evidence ideas were the focus and 'learning outcomes' of the lectures and were developed further with the students in the workshops. A characteristic of all the teaching was that the procedural ideas were introduced in deliberately simple contexts where the students were unlikely to be aware of a 'right answer' and where the cognitive demands of any substantive science ideas were low. This ensured that the focus was on the difficult procedural ideas. Since the students were all preparing for primary teaching such simple contexts were also relevant to their future professional practice. In the workshops, activities included:

- Whole class and small group discussion focused on the procedural ideas; such as how validity of design is affected by the values of variables selected, and how the validity of a relationship is affected by the range and interval of readings,
- Structured worksheet activities mainly in the form of question sheets; to enable students to practice and apply the procedural ideas in different simple contexts,
- Exploratory practicals to familiarise students with the procedural context; for instance, at the start of the module students were given equipment to build a beam bridge to explore the effect of the material on the bend when loaded, enabling students to recognise the 'thinking behind the doing',

- Practicals to illustrate specific procedural ideas; such as the effects of using different quality measuring instruments on the resultant data,
- Parts of investigations where students can apply some of the ideas they have learned; for example, students attempted to reduce the variation in repeated readings when students designed a way to measure the absorbency of a paper towel,
- A whole investigation; such as how the height a 'man' can 'jump' depends on the mass applied to a spring board, or whether the slide of a shoe is affected by the surface it is on.

**Instruments**

We used the same three instruments in both pre- and post-tests. They have been described extensively elsewhere (summarised in Gott and Roberts 2008) and are described briefly here. The instruments were presented in the student teachers' native language with the exception of the Evidence Test which was in English in both countries. Student teachers answered each instrument in their native language. The Turkish answers were translated into English by the second author and were subsequently checked by other experienced science educators.

We were interested in student teachers' understanding of the concepts of evidence which was assessed with short answer written items in the Evidence Test. We also wanted to see how the students synthesised and applied these ideas when carrying out a practical Open-ended Investigation, and to determine their ability to use the ideas to ask questions about a real socio-scientific issue in the Issue Task.

**Evidence Test:** This is a short answer written test, similar to those reported in Gott and Roberts (2004) and Roberts and Gott (2006) and is available from Gott and Roberts (2008). We have pre- and post-test data from 85 students in England and all 38 in Turkey. The test targets such things as measurement, experimental design and data presentation and analysis. The pre- and post- test, which took up to an hour for some students to complete, comprised items spanning the concepts of evidence. The pre-test was taken prior to any teaching and the post-test at the end of the module. Full details of the coding can be found at Gott and Roberts (2008).

**Open-ended Investigation:** Students were asked to investigate the problem 'Does the surface affect the slide of a shoe?' They were provided with access to any equipment they wanted in the laboratory and had to make their own decisions as to how to investigate the question. We have both pre- and post-test data from 73 students in England and 38 in Turkey. The students' written accounts were used as a means of assessing their ability to use their understanding about evidence to carry out an investigation, ('looking forward', Figure 2). We were also interested in the approaches they took to the problem. Full details of the coding employed for the English cohort can be found at Roberts, Gott, and Glaesser (2009). However, for this research a qualitative description of the student teachers' investigations are used. The pre-test was taken in the 1st session of the module, prior to any teaching and the post-test at the end of the module.

**Issue Task:** This instrument was measuring an aspect of scientific literacy; the ability to ask questions to scientists about a claim made in a local socio-scientific controversy (Roberts and Gott 2007; Gott and Roberts 2008). Students were presented with a claim by scientists of the safety of emissions from a cement works and were asked to write down any questions they would like to ask the scientists. We have pre-and post-test data from 65 students in England and 38 in Turkey. The pre-tests were taken before the module was taught. The post-tests were taken 7 weeks and 15 weeks after teaching in Turkey and England respectively. Full details of the coding can be found at Roberts and Gott (accepted). In this research we report the total number of questions asked, which we take as a measure of the student teachers' willingness to engage with the task; the number of questions about issues in the Broader Debate; and the questions about the evidence for the scientists' claim ('looking back', Figure 2) coded against the layers of the bull's-eye (Figure 1).

To summarise, we have described the concepts of evidence (summarised in Figure 1) which form the basis of the Evidence Module and are assessed in the Evidence Test. The Open-ended Investigation is represented as 'looking forward' in Figure 2, while the questions raised by the students in the Issue Task about the scientists' claim are coded against the layers in Figure 1 since they are 'looking back' from the claim (Figure 2).

## Results

We present a brief analysis of the pre-test data for each cohort in order to provide a description of the English and Turkish student teachers' understanding of evidence prior to the intervention. We note that there are many differences between the cohorts (Table 1) and differences in the pre-test data cannot be attributed to any one factor.

**Pre- Evidence Test results**

The results of the pre-Evidence Test are shown in Table 2.

[INSERT TABLE 2 ABOUT HERE]

Despite the differences in the samples, there is no significant difference in the total score between the English and Turkish student teachers. However, for one sub-set of questions, those about 'variables' (identifying variables and controlling variables), the English student teachers are significantly better than Turkish students. These are ideas fundamental to the design of experiments. The English cohort are likely to have experienced more practical work in their previous school science. Questions targetting data presentation and measurement showed no such difference between the cohorts.

**Pre- Open-ended Investigation results**

In the pre-Investigation, the majority of the English cohort were able to make at least some progress with the investigation. However, the written accounts were ritualised and demonstrated that they had little understanding of, or engagement with, the concepts of evidence per se but were falling back on routine linear approaches (Roberts, Gott, and Glaesser 2009). Despite having conducted investigations in their previous school science, the ability to tackle a novel task was very limited.

In contrast, the Turkish students, who had limited previous practical experience, struggled to make any progress with the pre-Investigation. The majority were unable even to get started and when faced with difficulties about deciding where to start, what to do and what to measure, they made very little progress. Their written accounts were so short that they could not be coded using the criteria previously used in England (Roberts, Gott, and Glaesser 2009).

**Pre- Issue Task results**

The results of the pre-Issue Task are shown in Table 3.

[INSERT TABLE 3 ABOUT HERE]

In the pre-Issue Task, the English student teachers asked significantly more questions than Turkish students, suggesting a greater willingness to question the scientists. In addition to asking more questions relevant to the broader debate, which included questions about the potential effect of the emissions on house prices, employment and the beauty of the surrounding area, they seem to have a realisation that a ‘claim’ can be interrogated. English students asked significantly more questions than Turkish students using ideas in layer B (data sets) and C (design of investigations) of the bull’s-eye. It is worth noting that neither cohort asked any questions using ideas in layer A (about the quality of a datum). There was no significant difference in questions using ideas in layers D, E or F (about comparisons with other data, the investigators themselves or factors which may have influenced stage of the investigations).

In summary, prior to the intervention the cohorts appeared to perform similarly overall on the Evidence Test with significant differences in questions about variables but the Turkish cohort really struggled with the practical context of the Open-ended Investigation and did not seem willing or able to ask questions in the Issue Task. Given the many differences between the cohorts we are unable to do more than speculate as to the reasons behind these differences.

We now turn to our main question. Will the Evidence Module bring about changes in the content knowledge of scientific evidence in these different cohorts? For each instrument we will present data comparing pre- and post-test data for each cohort to show the effect of the intervention. In addition we will also describe the relative effect of the Evidence Module on the two cohorts.

**Post- Evidence Test results**

As already described (Roberts and Gott 2007), very significant gains were made in all aspects of the English cohort’s understanding of evidence as measured by the Evidence Test (Table 4).

[INSERT TABLE 4 ABOUT HERE]

The Turkish cohort also significantly improved their understanding, as shown by the total score. However, there was no significant difference in questions focusing on data presentation and analysis (Table 5).

[INSERT TABLE 5 ABOUT HERE]

The Evidence Module has resulted in both cohorts, despite their differences, improving their understanding of evidence.

Table 6 is presented as a descriptive summary of the two cohorts’ post-Evidence Tests. Both cohorts achieved similar levels in the post-tests. We should not infer too much from this. It may be a reflection of the test construction.

[INSERT TABLE 6 ABOUT HERE]

**Post- Open-ended Investigation results**

The English cohort’s post-Open-ended Investigations are described more fully in Roberts, Gott, and Glaesser (2009). There was variation in their ability to carry out the investigation but all students demonstrated a much better application of the concepts of evidence. Instead of routinised ways of working, drawing on memories of similar activities, they were able to make decisions using understanding. The best students worked iteratively throughout the investigation, running trials and reflecting on the data as it was being collected to improve the validity and reliability of the evidence for their claim.

The Turkish cohort made considerable improvements compared with their pre-Investigation. However our data is incomplete since the style of the written account did not enable the same marking scheme to be utilised. Our observation is that compared to the pre-Investigations, the post-Investigations demonstrated that the Turkish students were now better able to carry out a practical investigation. Whether they have ‘caught up’ with English cohort is difficult to say because there is no data about their procedural understanding during the investigation, but they certainly improved in their ability to ‘have a go’.

**Post- Issue Task results**

The pre- and post- questions asked by the English cohort are shown in Table 7 and the Turkish cohort in Table 8.

[INSERT TABLES 7 & 8 ABOUT HERE]

The English cohort asked significantly fewer questions about the broader debate after teaching, focusing more on questions about the evidence for the claim. Of these, significant gains were made in questions about layer C (design and relationships) and layers E and F (factors that may have influenced the conduct of the investigation). By contrast, the significant changes in the Turkish cohort were with respect to questions in Layers B, C and D, albeit from very low pre-test scores. They also asked significantly more questions in total in the post-Issue Task. The interventions seem to have raised their awareness of the empirical basis for scientists’ claims.

Table 9 describes the questions asked in the post-Issue Task by both cohorts.

[INSERT TABLE 9 ABOUT HERE]

The Turkish student teachers asked significantly more questions than English students overall after teaching. This may reflect the seriousness with which the cohorts engaged in the task or inherent differences between the cohorts. The Turkish students asked significantly more questions about ideas in layers B, C, E and F. However, neither cohort showed a significant increase in questions using ideas in bull’s-eye layer A, which is disappointing.

## Conclusions and Implications

**Conclusion from Pre-tests**

Despite the different school experiences, both groups have similar levels of procedural understanding (as measured by the Evidence Test) but in questions about variables the English cohort understood more. The English student teachers asked more questions, including more evidence-based questions, about the scientists’ claim in the Issue Task. It may be that the English cohort was better able to recognise that scientists’ claims were based on empirical work. The biggest difference between the cohorts seems to be in their ability to even ‘have a go’ at the Investigation. The English student teachers demonstrated little by way of *understanding* what they were doing, but made limited progress by falling back on a ritualised, linear procedure. The Turkish student teachers floundered in this practical situation.

**Conclusion from Post-tests**

This replication study has demonstrated that the intervention, originally designed in the English context, has resulted in very significant increases in the content knowledge of the

Turkish cohort of student teachers, despite the cohorts having different characteristics. We have replicated the increases found in the English cohort in the Turkish cohort. The Evidence Module seems to be effective for both cohorts.

The intervention resulted in an increased understanding of the concepts of evidence in both cohorts. In the post-Evidence Test, overall they had similar levels of understanding as each. The qualitative observations of the Investigations indicate that both cohorts made progress from their pre-Investigations; the English cohort demonstrably increased their understanding and application of procedural ideas in their post-Investigations. The Turkish student teachers, while not working in such a sophisticated manner as the English, were far more willing to 'have a go' than they had been prior to the intervention. We can ascribe this difference to the intervention and can suggest that having an understanding of the ideas enables them to better make the many decisions necessary to make progress in an open-ended investigation. The differences between the cohorts in the post-Issue Task are interesting, with much greater increases after teaching shown by the Turkish cohort. In terms of an empowered form of scientific literacy, where students are willing to engage with the science in real socio-scientific issues, the intervention resulted in more procedural questions being raised although, even after the intervention disappointingly few students from either cohort asked questions about ideas at the centre of the bull's-eye, about the quality of each and every datum and its measurement. These ideas are important for any forensic examination of scientific evidence, since establishing the validity and reliability of the data are vital.

The data tentatively points to the Turkish cohort overall making relatively greater progress with the intervention than English student teachers, perhaps because brighter students are attracted to Primary Education programmes in Turkey. Overall this replication study indicates that the intervention has had a positive effect on both of these different cohorts. Their understanding of evidence, and its application while 'looking forward' in an investigation and 'looking back' in the Issue Task demonstrate a significantly improved understanding of the content knowledge relevant to teaching science curricula that have a focus on scientific evidence.

The intervention was approached from an 'understanding ideas about evidence' perspective. It involved the explicit teaching of the concepts of evidence suitably sequenced and structured, and involved targeted practical and non-practical activities. We conclude that the Evidence Module is a targeted, efficient intervention that enables student teachers, regardless of their previous educational experience, to develop a procedural understanding in a relatively short period of time. Such content knowledge about evidence is an understanding important for future primary science teachers.

(6021 words)

## References


American Association for the Advancement of Science. 1967. *Science - a process approach.* Washington DC: Ginn & Co.

Australian Curriculum, Assessment and Reporting Authority. 2009. *Shape of the Australian Curriculum: Science* http://www.acara.edu.au/verve/_resources/Australian_Curriculum_-_Science.pdf (accessed 3rd June 2010).

Bryce, T.G.K., J. McCall, J. Macgregor, I.J. Robertson, and R.A.J. Weston. 1983. *Techniques for the assessment of practical skills in foundation science.* London: Heinemann.

Buffler, A., S. Allie, and F. Lubben. 2001. The development of first year physics students' ideas about measurement in terms of point and set paradigms. *International Journal of Science Education* 23(11):1137–1156. doi:10.1080/09500690110039567.

Davis, B.C. 1989. *GASP: Graded assessment in science project.* London: Hutchinson.

DES and Welsh Office. 1989. *Science in the National Curriculum.* London: HMSO.

Duschl, R.A., H.A. Schweingruber, and A.W. Shouse, eds. 2006. *Taking science to school: Learning and teaching science in grades k-8.* Technical report, Committee on Science Learning, Kindergarten Through Eighth Grade. Board on Science Education, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Ekiz, D. 2006. Primary school teachers' attitudes towards educational research. *Educational Sciences: Theory and Practice.* 6(2):373-402.

Demirel, O. 2004. *Curriculum Development In Education From Theory To Practice* (in Turkish: *Kuramdan Uygulamaya Egitimde Program Gelistirme).* Ankara: Pegem A Publishing.

Glaesser, J., R. Gott, R. Roberts, and B. Cooper. 2009a. Underlying success in open-ended investigations in science: using qualitative comparative analysis to identify necessary and sufficient conditions. *Research in Science and Technology Education* 27(1):5-30.

Glaesser, J., R. Gott, R. Roberts, and B. Cooper. 2009b. The roles of substantive and procedural understanding in open-ended science investigations: Using fuzzy set Qualitative Comparative Analysis to compare two different tasks. *Research in Science Education* 39(4):595-624.

Gott, R. and S. Duggan. 2003. *Understanding and using scientific evidence.* London: Sage.

Gott, R. and S. Duggan. 2007. A framework for practical work in science and scientific literacy through argumentation. *Research in Science and Technological Education* 25(3):271–291.

Gott, R., S. Duggan, and R. Roberts. 2003. *The concepts of evidence.* http://www.dur.ac.uk/rosalyn.roberts/Evidence/cofev.htm (accessed 3rd June 2010).

Gott, R. and R. Roberts. 2004. A written test for procedural understanding: a way forward for assessment in the UK science curriculum? *Research in Science and Technological Education* **22**(1):5-21

Gott, R. and R. Roberts. 2008. *Concepts of evidence and their role in open-ended practical investigations and scientific literacy; background to published papers.* http://www.dur.ac.uk/resources/education/research/res_rep_short_master_final.pdf (accessed 3rd June 2010).

Grossman, P.L. 1990. *The making of a teacher: Teacher knowledge and teacher education.* New York: Teachers College Press.

Hodson, D. 1991. Practical work in science: time for a reappraisal. *Studies in Science Education* 19:175–184.

Ministry of Education, Culture and Science, Netherlands. Undated. http://english.minocw.nl/documenten/core%20objectives%20secondary%20education.pdf (accessed 3rd June 2010).

Ministry of National Education (Milli Egitim Bakanlıgı-MEB) 2000. *Primary Science Curriculum (grade 6, 7, 8).* Ankara: Directory of State Books.

Ministry of National Education (Milli Egitim Bakanlıgı-MEB) 2005. *Primary Science and Technology Curriculum (grade 6, 7, 8).* Ankara: Directory of State Books.

Polanyi, M. 1966. *The tacit dimension.* Gloucester, MA: Peter Smith.

Qualifications and Curriculum Authority. 2004. *Reforming science education for the 21st century; a commentary on the new GCSE criteria for awarding bodies.* London: QCA.

Roberts, R. 2009. How Science Works (HSW). *Education in Science* June 2009, no 233:30-31

Roberts, R. and R. Gott. 2003. Assessment of biology investigations. *Journal of Biological Education* 37(3):114-121.

Roberts, R. and R. Gott. 2006. Assessment of performance in practical science and pupil attributes. *Assessment in Education* 13(1):45-67

Roberts, R. and R. Gott. 2007. Questioning the Evidence: research to assess an aspect of scientific literacy. *Proceedings of European Science Education Research Association (ESERA) conference*. Malmo, Sweden, August 2007.

Roberts, R. and R. Gott. 2010. A Framework for Practical work, Argumentation and Scientific Literacy. In G. Cakmakci & M.F. Taşar (Eds.), *Contemporary science education research: scientific literacy and social aspects of science* (pp. 99-104). Ankara, Turkey: Pegem Akademi.

Roberts, R. and R. Gott. Accepted. Questioning the evidence for a claim in a socio-scientific issue: an aspect of scientific literacy. *Research in Science and Technological Education.*

Roberts, R., R. Gott, and J. Glaesser. 2010. Students' approaches to open-ended science investigation: the importance of substantive and procedural understanding. *Research Papers in Education* 25(4):377-407

Sahin-Pekmez, E. 2000. *Procedural understanding: Teachers' perceptions of Conceptual Basis of Practical Work*. University of Durham, School of Education, unpublished EdD Thesis.

Sahin-Pekmez, E. 2005. Science teachers' understanding of practical work (in Turkish: Fen Öğretmenlerinin laboratuvar çalışmaları ile ilgili görüşleri) *Journal of Buca Faculty of Education (in Turkish: Buca Egitim Fakultesi Dergisi)* 18:73-80.

Screen, P.A. 1986. The Warwick Process Science project. *The School Science Review* 72(260):17–24.

Shulman, L. 1986. Those who understand: Knowledge growth in teaching. *Educational Researcher* 15:4–14.

Shulman, L. 1987. Knowledge and teaching: Foundation of a new reform. *Harvard Review* 57:1–22.

Tytler, R., S. Duggan, and R. Gott. 2001a. Dimensions of evidence, the public understanding of science and science education. *International Journal of Science Education* 23(8):815–832. doi:10.1080/09500690010016058.

Tytler, R., S. Duggan, and R. Gott. 2001b. Public participation in an environmental dispute: implications for science education. *Public Understanding of Science* 10(4):343–364. doi:10.1088/0963-6625/10/4/301.

Unal, S., B. Costu, and F.O. Karatas. 2004. A general look at the science curriculum development studies in Turkey. *Journal of Gazi Faculty of Education* 24(2):183-202.

Figure 1: The bull's-eye diagram showing the summary 'layers' of the concepts of evidence (CofEv) (based on Roberts and Gott 2006). Essentially we need to be sure of the reliability and validity in each layer. Layers A to C focus on the ideas associated with data collection while ideas in the outer layers are important where the validity and reliability of the evidence may be affected by the broader context for the claim and its link to existing findings.

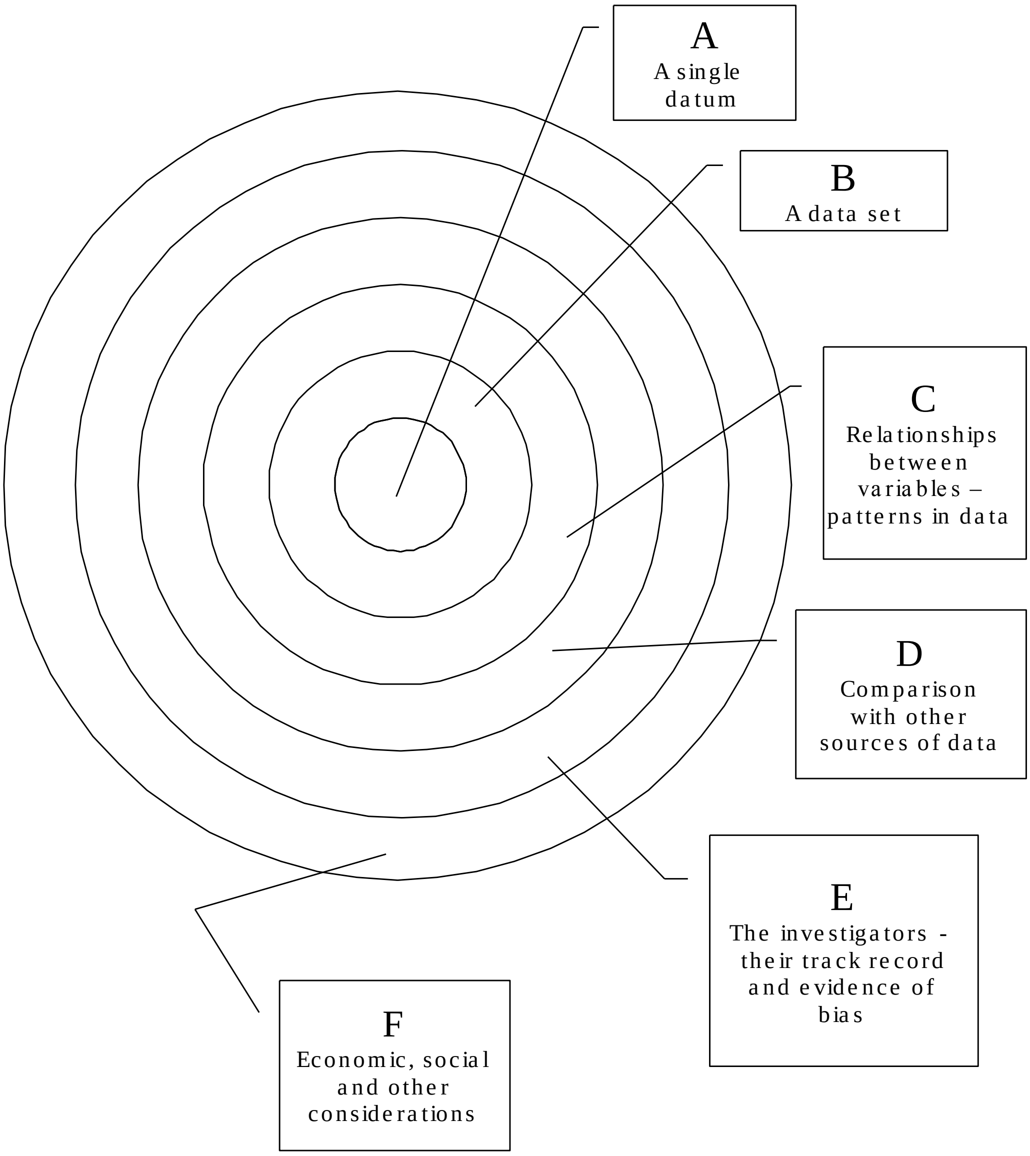

Figure 2: Investigations 'Looking forward' to make a claim and Evaluation 'looking back' from the claim.

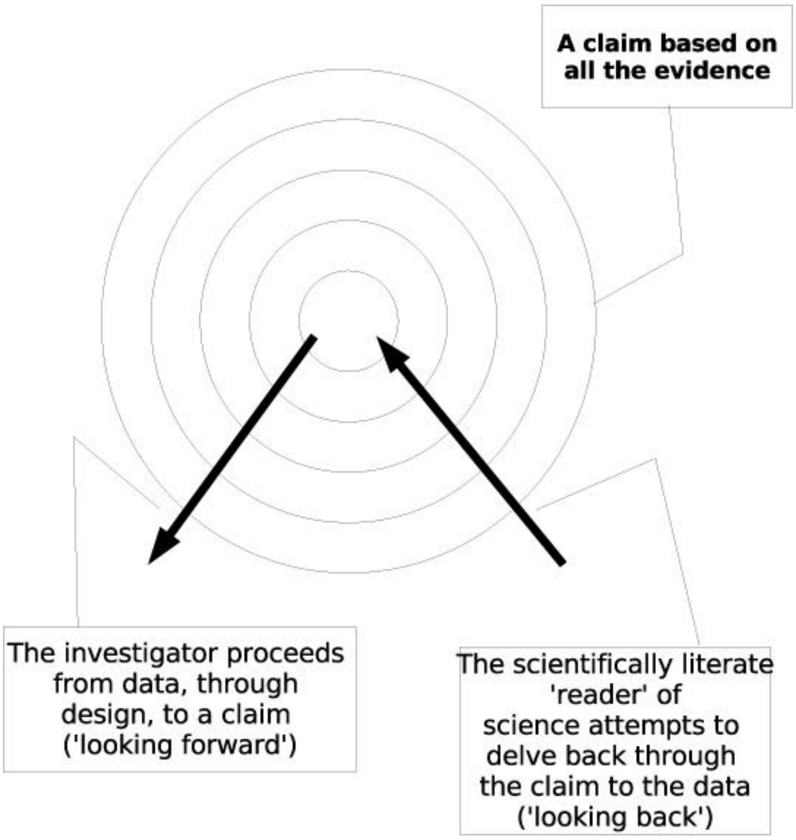

Table 1: The sample

| | **English cohort** | **Turkish cohort** |
|---|---|---|
| **n students** | 91 | 38 |
| **Mean age (years)** | 23.8 | 21.2 |
| **Age range (years)** | 19 - 45 | 20 - 26 |
| **Female : Male numbers** | 79 : 12 | 28 : 10 |
| **School science background** | Science compulsory until 16. Very few took any science when given the choice aged 16-18. Generally lacking confidence in science. The curriculum includes a procedural component and practical work is common. | Vast majority chose to specialise in science up to end of secondary schooling (aged 17). Traditional subject knowledge-focused curriculum. Relatively little practical work in school science. |
| **Entry qualifications compared with each University's other programmes.** | Ranked 27th out of 30 degree programmes. | Ranked 2nd out of 52 departments. |
| **English language level** | Native | Equivalent to IELTS of ~5.0 |

Table 2: Understandıng Evidence: Pre- Evidence Test results for each cohort

| | **English cohort**<br><br>**Mean %** | **Turkish cohort**<br><br>**Mean %** | ***t*-test significance**<br>*** sig at 0.05**<br>**** sig at 0.01** |
|---|---|---|---|
| **Total** | 52.9 | 46.9 | |
| **Sub-set of questions about data presentation and analysis** | 37.5 | 39.1 | |
| **Sub-set of questions about measurement** | 46.9 | 51.0 | |
| **Sub-set of questions about Variables** | 60.5 | 42.3 | ** |

Table 3: Questionsing Evidence: Pre- Issue Task results for each cohort

| | **English cohort**<br><br>**Mean number of questions** | **Turkish cohort**<br><br>**Mean number of questions** | ***t*-test significance**<br><br>*** sig at 0.05**<br>**** sig at 0.01** |
|---|---|---|---|
| **Total questions asked** | 7.5 | 3.1 | ** |
| **Questions about the Broader Debate** | 2.0 | 0.6 | ** |
| **Questions about the CofEv in layer A** | 0.0 | 0.0 | |
| **Questions about the CofEv in layer B** | 0.6 | 0.2 | ** |
| **Questions about the CofEv in layer C** | 1.1 | 0.5 | ** |
| **Questions about the CofEv in layer D** | 0.1 | 0.0 | |
| **Questions about the CofEv in layer E & F** | 0.5 | 0.3 | |

Table 4: Understanding Evidence: English cohort pre- and post-Evidence Test

| | **English cohort Pre-**<br><br>**Mean %** | **English cohort Post-**<br><br>**Mean %** | ***t*-test significance**<br><br>*** sig at 0.05**<br>**** sig at 0.01** |
|---|---|---|---|
| **Total** | 52.9 | 64.8 | ** |
| **Sub-set of questions about data presentation and analysis** | 37.5 | 46.7 | ** |
| **Sub-set of questions about measurement** | 46.9 | 64.7 | ** |
| **Sub-set of questions about Variables** | 60.5 | 72.4 | ** |

Table 5: Understanding Evidence: Turkish cohort pre- and post-Evidence Test

| | **Turkish cohort Pre-**<br><br>**Mean %** | **Turkish cohort Post-**<br><br>**Mean %** | ***t*-test significance**<br><br>*** sig at 0.05**<br>**** sig at 0.01** |
|---|---|---|---|
| **Total** | 46.9 | 66.2 | ** |
| **Sub-set of questions about data presentation and analysis** | 39.1 | 41.1 | |
| **Sub-set of questions about measurement** | 51.0 | 69.6 | ** |
| **Sub-set of questions about Variables** | 42.3 | 72.5 | ** |

Table 6: Understanding Evidence: English and Turkish cohorts, post-Evidence Test

| | **English cohort**<br><br>**Mean %** | **Turkish cohort**<br><br>**Mean %** | ***t*-test significance**<br><br>*** sig at 0.05**<br>**** sig at 0.01** |
|---|---|---|---|
| **Total** | 64.8 | 66.2 | |
| **Sub-set of questions about data presentation and analysis** | 46.7 | 41.1 | |
| **Sub-set of questions about measurement** | 64.7 | 69.6 | |
| **Sub-set of questions about Variables** | 72.4 | 72.5 | |

Table 7: Questioning Evidence: English cohort pre- and post- Issue Task

| | **English cohort Pre- Mean number of questions** | **English cohort Post- Mean number of questions** | ***t*-test significance** <br> *** sig at 0.05** <br> **** sig at 0.01** |
|---|---|---|---|
| **Total questions asked** | 7.5 | 7.9 | |
| **Questions about the Broader Debate** | 2.0 | 1.2 | ** |
| **Questions about the CofEv in layer A** | 0.0 | 0.03 | |
| **Questions about the CofEv in layer B** | 0.6 | 0.5 | |
| **Questions about the CofEv in layer C** | 1.1 | 1.7 | ** |
| **Questions about the CofEv in layer D** | 0.1 | 0.1 | |
| **Questions about the CofEv in layer E & F** | 0.5 | 0.9 | ** |

Table 8: Questioning Evidence: Turkish cohort pre- and post- Issue Task

| | **Turkish cohort Pre- Mean number of questions** | **Turkish cohort Post- Mean number of questions** | ***t*-test significance<br><br>* sig at 0.05<br>** sig at 0.01** |
|---|---|---|---|
| **Total questions asked** | 3.1 | 11.1 | ** |
| **Questions about the Broader Debate** | 0.6 | 0.7 | |
| **Questions about the CofEv in layer A** | 0.0 | 0.06 | |
| **Questions about the CofEv in layer B** | 0.2 | 1.2 | ** |
| **Questions about the CofEv in layer C** | 0.5 | 3.6 | ** |
| **Questions about the CofEv in layer D** | 0.0 | 0.2 | * |
| **Questions about the CofEv in layer E & F** | 0.3 | 0.5 | |

Table 9: Questioning Evidence: English and Turkish cohorts, post- Issue Task

| | **English cohort**<br><br>**Mean number of questions** | **Turkish cohort**<br><br>**Mean number of questions** | ***t*-test significance**<br><br>*** sig at 0.05**<br>**** sig at 0.01** |
|---|---|---|---|
| **Total questions asked** | 7.9 | 11.1 | ** |
| **Questions about the Broader Debate** | 1.2 | 0.7 | * |
| **Questions about the CofEv in layer A** | 0.03 | 0.06 | |
| **Questions about the CofEv in layer B** | 0.5 | 1.2 | ** |
| **Questions about the CofEv in layer C** | 1.7 | 3.6 | ** |
| **Questions about the CofEv in layer D** | 0.1 | 0.2 | |
| **Questions about the CofEv in layer E & F** | 0.9 | 0.5 | * |